

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Metrics for pitch collections

### Conference or Workshop Item

#### How to cite:

Milne, Andrew J.; Sethares, William A.; Laney, Robin and Sharp, David B. (2010). Metrics for pitch collections. In: Proceedings of the International Conference of Music Perception and Cognition 2010 (ICMPC 11), 23-27 Aug 2010, Seattle, USA.

For guidance on citations see [FAQs](#).

© 2010 Andrew Milne

Version: Version of Record

Link(s) to article on publisher's website:  
<http://depts.washington.edu/icmpc11/>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# METRICS FOR PITCH COLLECTIONS

*Andrew J. Milne<sup>1</sup>, William A. Sethares<sup>2</sup>, Robin Laney<sup>1</sup>, David B. Sharp<sup>1</sup>*

<sup>1</sup>The Open University, UK

<sup>2</sup>University of Wisconsin-Madison, USA

## ABSTRACT

Models of the perceived distance between pairs of pitch collections are a core component of broader models of the perception of tonality as a whole. Numerous different distance measures have been proposed, including voice-leading, psychoacoustic, and pitch and interval class distances; but, so far, there has been no attempt to bind these different measures into a single mathematical framework, nor to incorporate the uncertain or probabilistic nature of pitch perception (whereby tones with similar frequencies may, or may not, be heard as having the same pitch).

To achieve these aims, we embed pitch collections in novel multi-way expectation arrays, and show how metrics between such arrays can model the perceived dissimilarity of the pitch collections they embed. By modeling the uncertainties of human pitch perception, expectation arrays indicate the expected number of tones, ordered pairs of tones, ordered triples of tones and so forth, that are heard as having any given pitch, dyad of pitches, triad of pitches, and so forth. The pitches can be either absolute or relative (in which case the arrays are invariant with respect to transposition).

We provide a number of examples that show how the metrics accord well with musical intuition, and suggest some ways in which this work may be developed.

## 1. BACKGROUND AND AIMS

A pitch collection may comprise the pitches of tones in a chord, a scale, a tuning, or the virtual and spectral pitches heard in response to complex tones or chords. Modeling the perceived distance (the similarity or dissimilarity) between pairs of pitch collections has a number of important applications in music analysis and composition, in modeling of musical cognition, and in the design of musical tunings. For example, voice-leading distances model the overall distance between two chords as a function of the pitch distance moved by each voice (see Tymoczko (2006) for a survey); musical set theory considers the similarities between the interval (or triad, tetrad, etc.) contents of pitch collections (see Castrén (1994) for a survey); psychoacoustic models of chordal distance (Parncutt, 1989; Milne, 2009) have treated tones or chords as collections of virtual and spectral pitches (Terhardt, Stoll, & Seewann, 1982; Zwicker & Fastl, 1999) to determine their affinity; tuning theory requires measures that can determine the distance between scale tunings and, notably, the extent to which different scale tunings can approximate privileged tunings of intervals or chords (e.g., just intonation intervals with frequency ratios such as 3/2 and 5/4, or chords with frequency ratios such as 4:5:6:7).

## 2. METHOD

### 2.1 Pitch (Class) Vectors

A pitch (class) vector contains elements whose values indicate pitches (typically in semitones). Standard metrics between two such vectors are based only on the pitch distances between elements in matching positions in the two vectors. For this reason, such pitch metrics are only meaningful when each tone in one pitch collection has a privileged relationship with a unique tone in another pitch collection; for example, when each element represents a different voice (bass, tenor, alto, soprano), or scale degree, or even metrical or ordinal position in a melody.

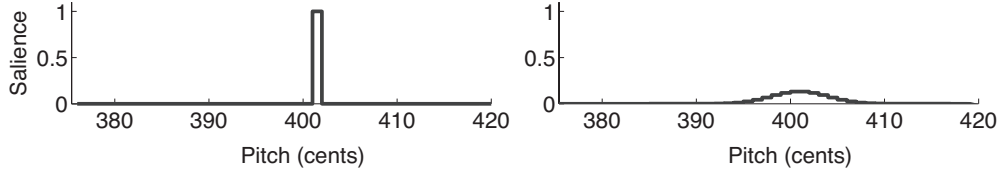
Such metrics are not suitable when the pitches cannot be uniquely categorized in the manner described above. This occurs when two collections have a different number of pitches, or when there are so many pitches that categorization is inappropriate. For example, when modeling the distance between the large sets of spectral or virtual pitches heard in response to complex tones or chords, there is no unique way to reasonably align each spectral pitch of one complex tone or chord with each spectral pitch of another (Sethares, Milne, Tiedje, Prechtel, & Plamondon, 2009) and, even if there were, it is not realistic to expect humans to track the “movements” of such a multitude of pitches.

### 2.2 Expectation Arrays

We present a novel family of pitch embeddings (*expectation arrays*), and associated metrics, that can be applied to the above areas. Expectation arrays model the uncertainties of pitch perception by “smearing” each pitch over a range of possible values, and the width of the smearing can be derived directly from experimentally determined frequency difference limens (Moore, Glasberg, & Shailer, 1984; Roederer, 1994). The effect of this pitch smearing is significant whenever pitches in one collection are similar, but non-identical, to pitches in another; for example, when comparing scales with different microtonal tunings, or the collections of virtual and spectral pitches produced in response to different chords.

The arrays can embed either absolute or relative pitches (denoted *absolute* and *relative expectation arrays*, respectively): in the latter case, embeddings of pitch collections that differ only by transposition have zero distance; a useful feature that relates similarity to structure.

Depending on their number of dimensions, expectation arrays indicate the expected number of tones, ordered pairs of tones, ordered triples of tones, and so forth, that will be heard as having any given pitch, dyad of pitches, triad of pitches, and so forth. This enables different pitch collections to be compared according



**Figure 1:** The effect of probabilistic smoothing on pitch domain embeddings.

to their monad (single pitch), dyad, triad, tetrad, and so forth, content. To see why such comparisons are significant, consider a simple example using major and minor triads: These contain the same set of intervals (and hence they have zero dyadic distance) but these intervals are arranged in different ways (and hence have non-zero triadic distance). Thus the two types of embedding may capture the way major and minor triads are heard to be simultaneously similar and different.

### 2.3 Derivation of Expectation Arrays

Expectation arrays are derived by transforming each element of a pitch vector into a characteristic (indicator) function in the pitch domain (as shown in left side of Figure 1). These vectors are then weighted by their *saliency* (probability of being heard (Parncutt, 1989)) and convolved with a probability mass function, which “smears” the spikes over a range of pitch values to model perceptual uncertainty (see the right side of Figure 1).

The saliency value of tone  $i$  at pitch  $j$  is denoted  $x_{\text{pcr}_{i,j}}$  and these values are used to calculate the family of expectation arrays detailed below.

If it is assumed that the hearing of any tone does not affect the probability of hearing any other tone, the expected number of tones that will be heard as having pitch (class)  $j$  is given by the *absolute monad expectation array*  $\hat{\mathbf{X}}_e^{(1)}$ , which contains elements

$$x_{e_j} = \sum_{i=1}^d x_{\text{pcr}_{i,j}},$$

where  $d$  is the number of tones.

The expected number of tones that will be heard, regardless of pitch, is given by summing the absolute monad expectation array over  $j$  to give the *relative monad expectation array*  $\hat{\mathbf{X}}_e^{(0)}$  (this “array” is actually a scalar)

$$\hat{x}_e = \sum_{j=0}^{q-1} x_{e_j}$$

The expected number of ordered pairs of tones that will be heard as having the dyad of pitches  $j$  and  $j + k_2$  is given by the *absolute dyad expectation array*  $\hat{\mathbf{X}}_e^{(2)}$ , which contains elements

$$x_{e_{j,k_2}} = \sum_{\substack{(i_1,i_2) \in D^2: \\ i_1 \neq i_2}} x_{\text{pcr}_{i_1,j}} x_{\text{pcr}_{i_2,j+k_2}},$$

where  $D = \{1, 2, \dots, d\}$ .

The expected number of ordered pairs of tones that will be heard as having the interval  $k_2$  is given by summing the absolute dyad expectation array over  $j$  to give the *relative dyad expectation array*  $\hat{\mathbf{X}}_e^{(1)}$ , which contains elements

$$\hat{x}_{e_{k_2}} = \sum_j x_{e_{j,k_2}}$$

This process can be continued for triads of pitches, tetrads of pitches, pentads of pitches, and so forth. A generalized form for *absolute r-ad expectation arrays* can be written accordingly:

$$x_{e_{j,k_2,\dots,k_r}} = \sum_{\substack{(i_1,\dots,i_r) \in D^r: \\ i_n \neq i_o}} \prod_{m=1}^r x_{\text{pcr}_{i_m,j+k_m}}$$

where  $k_1 = 0$ . *Relative r-ad expectation arrays* are given by summing over  $j$ .

The time taken to calculate the arrays can be substantially decreased by using algebraic manipulations, only calculating for non-zero values, and taking advantage of symmetries in the resulting array to calculate only those values that are not duplicated. (A full description of these methods is beyond the scope of this short paper).

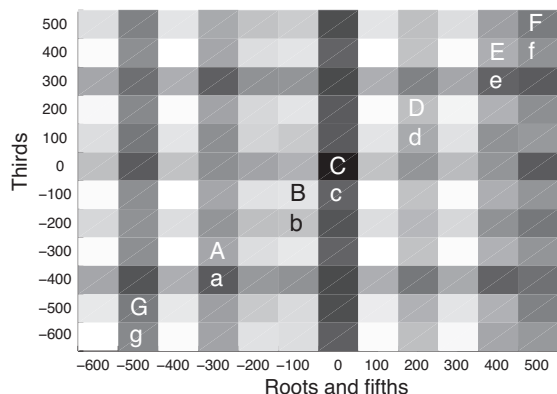
## 3. APPLICATIONS

We present here a few examples of applications of the expectation arrays.

### 3.1 Tonal affinity

Figure 2 shows the *spectral distance* between the spectral pitches (first ten harmonics of each tone) of a reference major triad and all 12-tone equal temperament triads that contain a perfect fifth. The spectral distance is calculated using absolute monad expectation arrays and a cosine metric. The horizontal axis shows the pitch distance from the reference triad’s root and fifth, the vertical axis

shows the pitch distance from the reference triad's third. All root-position major and minor triads lie on the central diagonal and some of these have been labeled for reference. The darker the grey the smaller the spectral distance from the reference triad (i.e., the greater the spectral affinity).

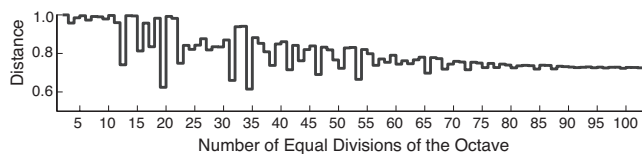


**Figure 2:** Spectral distances between a reference C-major triad and all possible three-voice triads with a perfect fifth.

This model suggests that the triad pair {C-major, d-minor} has greater spectral affinity (lower distance) than the neighboring triad pair {C-major, D-major}; the triad pair {C-major, F-major} has greater spectral affinity than the neighboring triad pair {C-major, F#-major}; the triad pair {C-major, e-minor} has greater spectral affinity than the neighboring triad pair {C-major, E-major}; and so forth. These results seem indicative of the tonal function of these triad pairings: the latter pair in each case is typically heard as requiring resolution, the former pair in each case is not (Milne, 2009). This suggests that such metrics may provide effective models for the feelings of expectation and resolution induced by successions of chords in tonal-harmonic music.

### 3.2 Tuning systems

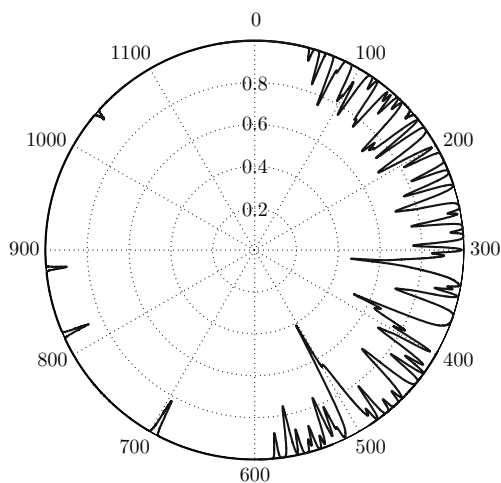
Figure 3 shows the distance (using relative dyad expectation arrays and a cosine metric) between all equal temperaments from 2 steps per octave to 102 steps per octave.



**Figure 3:** The distance between embeddings of a just intonation major triad and equal divisions of the octave.

It is interesting to observe that the distances approach a flat line where increasing the number of divisions of the octave is no longer beneficial, and that the most prominent minima fall at the familiar 12-tone equal temperament and at other alternative equal tunings (e.g., 19-, 22-, 31-, 34-, 41-, 46-, and 53-edo) that are well-known in the microtonal literature.

Figure 4 shows the distances between the relative expectation embeddings (dyadic on the right, triadic on the left) of a just intonation major triad {0, 386.3, 702} and the seven-tone scales generated by an interval whose size is incremented from 0 cents to 1199.9 cents (in 0.1 cent increments). (Note that if the same distance measure is used for the whole tuning range, the chart is symmetrical about the line connecting 0 and 600 cents.)



**Figure 4:** Distances (dyadic on the right, triadic on the left) between a just intonation major triad {0, 386.3, 702} and the seven-tone scales generated by an interval whose size is incremented from 0 cents to 1199.9 cents.

It is interesting to observe that for low-cardinality generated scales (such as this seven-tone scale) there are many tunings that provide a large number of good approximations to the intervals in the just intonation triad (the perfect fifth and major and minor thirds and their inversions), but only a few tunings provide a large number of good approximations to a complete major triad (and notably the best, in this model, is the familiar meantone tuning of approximately 696 cents).

## 4. DISCUSSION

We have presented a novel family of embeddings and metrics for determining the distance between pitch collections. The embeddings are based upon psychoacoustic principles (through the use of Gaussian smoothing) and may be useful as components in broader models of the perception and cognition of music. Indeed, to model any specific aspect of musical perception, a variety of

appropriate embeddings may be linearly combined, with their weightings, the weightings of the tone saliences (if appropriate), and the type of metric, as free parameters to be determined from experimental data.

We have focused on expectation arrays, but the underlying pitch (class) response matrices can also be used to generate salience arrays, which give the probability of hearing any given  $r$ -ad of pitches. There may also be scope in applying Fourier transforms to the embeddings in order to determine similarities in the spectrum of equal temperaments that approximate various pitch collections.

The methods are also applicable to any domain involving the perception of discrete stimuli. An obvious example is the perception of timing in rhythms, with time replacing pitch so the smoothing represents perceptual or cognitive inaccuracies in timing; for example, it might be possible to embed a rhythmic motif containing four events in a relative tetrad expectation matrix (in the time domain), and compare this with a selection of other similarly embedded rhythm patterns to find one with the closest match (i.e., one that contains the greatest number of patterns that are similar to the complete motif).

## 5. REFERENCES

- Castrén, M. (1994). *RECREL: A Similarity Measure for Set-classes*. Helsinki: Sibelius Academy.
- Milne, A. J. (2009). *A psychoacoustic model of harmonic cadences*. Master's Thesis, University of Jyväskylä, Department of Music, Jyväskylä.
- Moore, B. C., Glasberg, B. R., & Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics with complex tones. *The Journal of the Acoustical Society of America*, 75 (2), 500–561.
- Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.
- Roederer, J. G. (1994). *The Physics and Psychophysics of Music*. New York: Springer-Verlag.
- Sethares, W. A., Milne, A. J., Tiedje, S., Prechtl, A., & Plamondon, J. (2009). Spectral tools for Dynamic Tonality and audio morphing. *Computer Music Journal*, 33 (2).
- Terhardt, E., Stoll, G., & Seewann, M. (1982). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *Journal of the Acoustical Society of America*, 71 (3), 671–678.
- Tymoczko, D. (2006). Supporting online material for the geometry of musical chords. *Science*, 313 (72).
- Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and Models*. Berlin: Springer.